-NoValue-

Bioconductor's Computational Ecosystem for Genomic Data Science in Cancer

Name of First Author[0000−1111−2222−3333] and
Name of Second Author[1111−2222−3333−4444]

Abstract The Bioconductor project enters its third decade with over two thousand packages for genomic data science, over 100,000 annotation and experiment resources, and a global system for convenient distribution to researchers. Over 60,000 PubMed Central citations and terabytes of content shipped per month attest to the impact of the project on cancer genomic data science. This report provides an overview of cancer genomics resources in Bioconductor. After an overview of Bioconductor project principles, we address exploration of institutionally curated cancer genomics data such as TCGA. We then review genomic annotation and ontology resources relevant to cancer and then briefly survey analytical workflows addressing specific topics in cancer genomics. Concluding sections cover how new software and data resources are brought into the ecosystem and how the project is tackling needs for training of the research workforce. Bioconductor's strategies for supporting methods developers and researchers in cancer genomics are evolving along with experimental and computational technologies. All the tools described in this report are backed by regularly maintained learning resources that can be used locally or in cloud computing environments.

1 Introduction

Computation is a central component of cancer genomics research. Tumor sequencing is the basis of computational investigation of mutational, epigenetic and immunologic processes associated with cancer initiation and progression. Numerous computational workflows have been produced to profile tumor cell transcriptomes

Name of First Author
Name, Address of Institute, e-mail: name@email.address

Name of Second Author
Name, Address of Institute e-mail: name@email.address

and proteomes. New technologies promise to unite sequence-based characterizations with digital histopathology, ultimately driving efforts in molecule design and evaluation to produce patient-centered treatments.

Bioconductor is an open source software project with a 20 year history of uniting biostatisticians, bioinformaticians, and genome researchers in the creation of an ecosystem of data, annotation, and analysis resources for research in genome-scale biology. This paper will review current approaches of the project to advancing cancer genomics. After a brief discussion of basic principles of the Bioconductor project, we will present a "top down" survey of resources useful for cancer bioinformatics. Primary sections address

- how to explore institutionally curated cancer genomics data
- genomic annotation resources relevant to cancer genomics
- analytical workflows
- components for introducing new data or analyses
- pedagogics and workforce development.

2 Bioconductor principles

2.1 R packages and vignettes

Software tools and data resources in Bioconductor are organized into "R packages". These are collections of folders with data, code (principally R functions), and documentation following a protocol specified in https://cran.r-project.org/doc/manuals/R-exts.htmlWriting R Extensions. R packages have a DESCRIPTION file with metadata about package contents and provenance. Package structure can be checked for validity using the `R CMD check` facility. Documentation of code and data can be programmatically checked for existence and validity. The DESCRIPTION file for a package specifies its version and also gives precise definition of how an R package may depend upon versions of other packages.

At its inception, Bioconductor introduced a new approach to holistic package documentation called "vignette". Vignettes provide narrative and explanation interleaved with executable code describing package operations. While R function manual pages describe the operation of individual functions, vignettes illustrate the interoperation of package components and provide motivation for both package design but also context for its use.

2.2 R package repositories; repository evolution

Bioconductor software forms a coherent ecosystem that can be checked for consistency of versions of all packages available in a given installation of R. Bioconductor

packages may specify dependency on other Bioconductor packages, or packages that are available in the CRAN repository. Bioconductor does not include packages with dependencies on "github-only" packages. Later in this paper we will provide details on package quality assurance that provide a rationale for this restriction.

Major updates to the R language occur annually, and updates are preceded by careful assessment of effects of language change on Bioconductor package operations. These effects can be identified through changes in the output of R CMD check. The Bioconductor ecosystem is updated twice a year, once to coincide with update to R, and once about six months later. The semianual updates reflect the need to track developments in the fast-moving field of genomic data science.

2.3 Package quality assessment; installation consistency

The BiocCheck function is used to provide more stringent assessment of package compliance with basic principles of the Bioconductor ecosystem.

The BiocManager package provides for installing and updating package and has functionality for verifying the coherence and version status of the currently installed package collection. This is important in the context of a language and package ecosystem that changes every six months, while analyses may take years to complete. Tools for recreating past package collections are available to assist in reproducing outputs of prior analyses.

2.4 Unifying assay and sample data: SummarizedExperiment and MultiAssayExperiment

Most of the data from genome-scale experiments to be discussed in this chapter are organized in special data containers rooted in the concepts of the SummarizedExperiment class. Briefly, assay data are thought of as occupying a $G \times N$ array, and sample level data occupy an $N \times K$ table. The array and the table are linked together in the SummarizedExperiment; see Figure 1.

Multiple representations of assay results may be managed in this structure, but all assay arrays must have dimensions $G \times N$.

For experiment collections in which the same samples are subjected to multiple genome-scale assays, MultiAssayExperiment containers are used. See Figure 2 for the layout.

Further details on these data structures will be provided in section **??**.

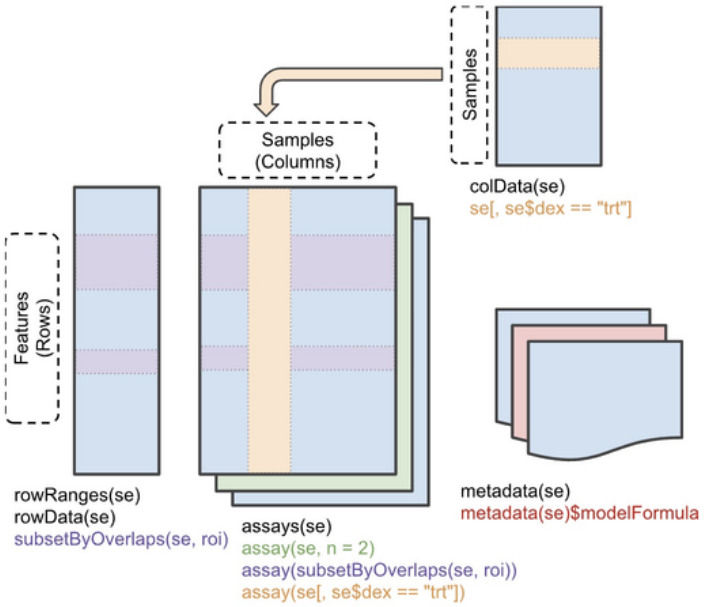

Fig. 1 SummarizedExperiment schematic.

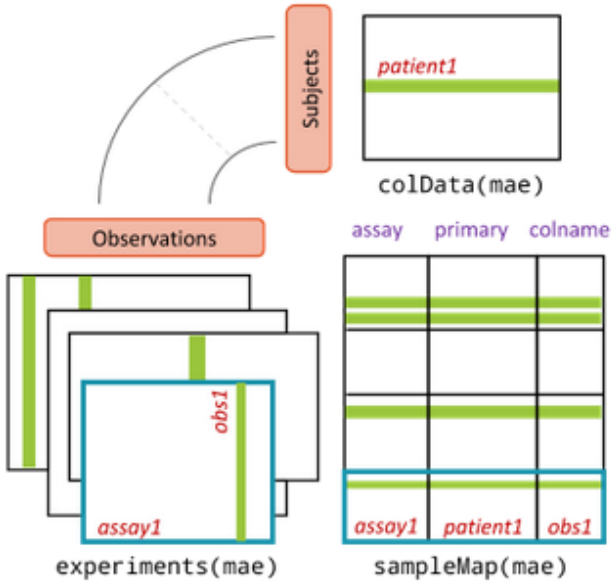

Fig. 2 MultiAssayExperiment schematic.

2.5 Downloading and caching cancer genomics data and annotations

Downloading and managing data from various online resources can be excessively time consuming. Bioconductor encourages data caching for increased efficiency and reproducibility. The caching data methods employed in Bioconductor allow analysis code to concisely refer to data resources as needed, with minimal attention to how data are stored, retrieved or transformed. It allows for easy management and reuse of data that are on remote servers or in cloud, storing source location and providing information for data updates. The BiocFileCache Bioconductor package handles data management from within R.

BiocFileCache is a general-use caching system but Bioconductor also provides "Hubs", AnnotationHub and ExperimentHub, to help distributed annotation or experimental data hosted externally. Both AnnotationHub and ExperimentHub use BiocFileCache to handle download and caching of data.

AnnotationHub provides a centralized repository of diverse genomic annotations, facilitating easy access and integration into analyses. Researchers can seamlessly retrieve information such as genomic features, functional annotations, and variant data, streamlining the annotation process for their analyses.

ExperimentHub extends this concept to experimental data. It serves as a centralized hub for storing and sharing curated experiment-level datasets, allowing researchers to access a wide range of experimental designs and conditions. This cloud-based infrastructure enhances collaboration and promotes the reproducibility of analyses across different laboratories.

The curatedTCGAData package provides some resources through ExperimentHub, as do many other self-identified "CancerData" resources. Once the ExperimentHub is loaded, it can be queried for terms of interest.

```
library(ExperimentHub)
eh = ExperimentHub()
query(eh, "CancerData")
```

Multiple terms can be used to narrow results before choosing a download.

```
query(eh, c("CancerData", "esophageal"))
# ExperimentHub with 2 records}
# snapshotDate(): 2023-10-24}
# $dataprovider: University of California San Francisco}
# $species: Homo sapiens}
# $rdataclass: RangedSummarizedExperiment, data.frame}
# additional mcols(): taxonomyid, genome, description,}
#   coordinate_1_based, maintainer, rdatadateadded, preparerclass, tags,}
#   rdatapath, sourceurl, sourcetype }
# retrieve records with, e.g., object[["EH8527"]]
#            title
#   EH8527 | cao_esophageal_wgbs_hg19
#   EH8530 | cao_esophageal_transcript_counts
```

Similarly AnnotationHub files can be downloaded for annotating data. For example, the ensembl 110 release of gene and protein annotations are obtained with the following:

```
library(AnnotationHub)
ah = AnnotationHub()
query(ah, c("ensembl", "110"))
```

3 Section Heading

Use the template *chapter.tex* together with the document class SVMono (monograph-type books) or SVMult (edited books) to style the various elements of your chapter content.

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations. And please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

4 Section Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Use the standard `equation` environment to typeset your equations, e.g.

$$a \times b = c \ , \tag{1}$$

however, for multiline equations we recommend to use the `eqnarray` environment[1].

$$\left|\nabla U_\alpha^\mu(y)\right| \le \frac{1}{d-\alpha} \int \left|\nabla \frac{1}{|\xi - y|^{d-\alpha}}\right| d\mu(\xi) = \int \frac{1}{|\xi - y|^{d-\alpha+1}}\, d\mu(\xi) \tag{2}$$

$$= (d-\alpha+1) \int\limits_{d(y)}^{\infty} \frac{\mu(B(y,r))}{r^{d-\alpha+2}}\, dr \le (d-\alpha+1) \int\limits_{d(y)}^{\infty} \frac{r^{d-\alpha}}{r^{d-\alpha+2}}\, dr \tag{3}$$

[1] In physics texts please activate the class option `vecphys` to depict your vectors in ***boldface-italic*** type - as is customary for a wide range of physical subjects

4.1 Subsection Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

> Please do not use quotation marks when quoting texts! Simply use the `quotation` environment – it will automatically be rendered in line with the preferred layout.

4.1.1 Subsubsection Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.1, see also Fig. 3[2]

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Paragraph Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

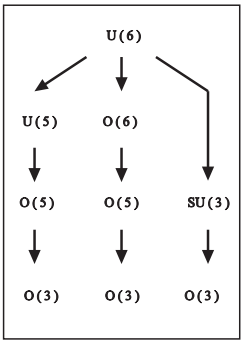

Fig. 3 If the width of the figure is less than 7.8 cm use the `sidecapion` command to flush the caption on the left side of the page. If the figure is positioned at the top of the page, align the sidecaption with the top of the figure – to achieve this you simply need to use the optional argument `[t]` with the `sidecaption` command

[2] If you copy text passages, figures, or tables from other works, you must obtain *permission* from the copyright holder (usually the original publisher). Please enclose the signed permission with the manuscript. The sources must be acknowledged either in the captions, as footnotes or in a separate section of the book.

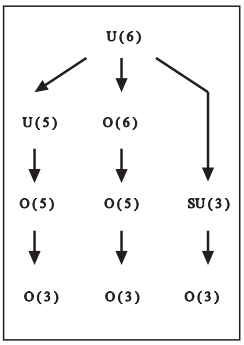

Fig. 4 If the width of the figure is less than 7.8 cm use the `sidecapion` command to flush the caption on the left side of the page. If the figure is positioned at the top of the page, align the sidecaption with the top of the figure – to achieve this you simply need to use the optional argument `[t]` with the `sidecaption` command

For typesetting numbered lists we recommend to use the `enumerate` environment – it will automatically rendered in line with the preferred layout.

1. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
   a. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
   b. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
2. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.

Subparagraph Heading

In order to avoid simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4, see also Fig. 4.

For unnumbered list we recommend to use the `itemize` environment – it will automatically be rendered in line with the preferred layout.

- Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development, cf. Table 1.
  – Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
  – Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
- Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.

Table 1 Please write your table caption here

Classes	Subclass	Length	Action Mechanism
Translation	mRNA[a]	22 (19–25)	Translation repression, mRNA cleavage
Translation	mRNA cleavage	21	mRNA cleavage
Translation	mRNA	21–22	mRNA cleavage
Translation	mRNA	24–26	Histone and DNA Modification

[a] Table foot note (with superscript)

Run-in Heading Boldface Version Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

Run-in Heading Boldface and Italic Version Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

Run-in Heading Displayed Version
Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

5 Section Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

If you want to list definitions or the like we recommend to use the enhanced `description` environment – it will automatically rendered in line with the preferred layout.

Type 1 That addresses central themes pertainng to migration, health, and disease. In Sect. 3, Wilson discusses the role of human migration in infectious disease distributions and patterns.

Type 2 That addresses central themes pertainng to migration, health, and disease. In Sect. 4.1, Wilson discusses the role of human migration in infectious disease distributions and patterns.

5.1 Subsection Heading

In order to avoid simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Use the LaTeX automatism

for all your cross-references and citations citations as has already been described in Sect. 4.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

If you want to emphasize complete paragraphs of texts we recommend to use the newly defined class option `graybox` and the newly defined environment `svgraybox`. This will produce a 15 percent screened box 'behind' your text.

If you want to emphasize complete paragraphs of texts we recommend to use the newly defined class option and environment `svgraybox`. This will produce a 15 percent screened box 'behind' your text.

5.1.1 Subsubsection Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Theorem 1 *Theorem text goes here.*

Definition 1 Definition text goes here.

Proof. Proof text goes here. □

Paragraph Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 4.

Note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Theorem 2 *Theorem text goes here.*

Definition 2 Definition text goes here.

Proof. Proof text goes here. □

Trailer Head

If you want to emphasize complete paragraphs of texts in an `Trailer Head` we recommend to use

```
\begin{trailer}{Trailer Head}
...
\end{trailer}
```

? Questions

If you want to emphasize complete paragraphs of texts in an `Questions` we recommend to use

```
\begin{questype}{Questions}
...
\end{questype}
```

> Important

If you want to emphasize complete paragraphs of texts in an `Important` we recommend to use

```
\begin{important}{Important}
...
\end{important}
```

! Attention

If you want to emphasize complete paragraphs of texts in an `Attention` we recommend to use

```
\begin{warning}{Attention}
...
\end{warning}
```

Program Code

If you want to emphasize complete paragraphs of texts in an `Program Code` we recommend to use

```
\begin{programcode}{Program Code}
\begin{verbatim}...\end{verbatim}
\end{programcode}
```

Tips

If you want to emphasize complete paragraphs of texts in an `Tips` we recommend to use

```
\begin{tips}{Tips}
...
\end{tips}
```

Overview

If you want to emphasize complete paragraphs of texts in an `Overview` we recommend to use

```
\begin{overview}{Overview}
...
\end{overview}
```

Background Information

If you want to emphasize complete paragraphs of texts in an `Background Information` we recommend to use

```
\begin{backgroundinformation}{Background Information}
...
\end{backgroundinformation}
```

Legal Text

If you want to emphasize complete paragraphs of texts in an `Legal Text` we recommend to use

```
\begin{legaltext}{Legal Text}
...
\end{legaltext}
```

Acknowledgements If you want to include acknowledgments of assistance and the like at the end of an individual chapter please use the `acknowledgement` environment – it will automatically be rendered in line with the preferred layout.

Competing Interests Please declare any competing interests in the context of your chapter. The following sentences can be regarded as examples.
This study was funded by [X] [grant number X].
[Author A] has a received research grant from [Company W].
[Author B] has received a speaker honorarium from [Company X] and owns stock in [Company Y].
[Author C] is a member of [committee Z].
The authors have no conflicts of interest to declare that are relevant to the content of this chapter.

Ethics Approval If your chapter includes primary studies with humans please declare the adherence of ethical standards. Example text: This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of University B (Date.../No. ...).
In addition, for human participants, authors are required to include a statement that informed consent (to participate and/or to publish) was obtained from individual participants or parents/guardians if the participant is minor or incapable.
If animals are studied, authors should make sure that the legal requirements or guidelines in the country and/or state or province for the care and use of animals have been followed or specify that no ethics approval was required.

Appendix

When placed at the end of a chapter or contribution (as opposed to at the end of the book), the numbering of tables, figures, and equations in the appendix section continues on from that in the main text. Hence please *do not* use the `appendix` command when writing an appendix at the end of your chapter or contribution. If there is only one the appendix is designated "Appendix", or "Appendix 1", or "Appendix 2", etc. if there is more than one.

[1] is our latest.

$$a \times b = c \tag{4}$$

References

1. Aevermann, Brian D., Mark Novotny, Trygve Bakken, Jeremy A. Miller, Alexander D. Diehl, David Osumi-Sutherland, Roger S. Lasken, Ed S. Lein, and Richard H. Scheuermann. 2018. "Cell Type Discovery Using Single-Cell Transcriptomics: Implications for Ontological Representation." *Human Molecular Genetics* 27 (R1): R40–R47. https://doi.org/10.1093/hmg/ddy100.
2. Aganezov, Sergey, Stephanie M. Yan, Daniela C. Soto, Melanie Kirsche, Samantha Zarate, Pavel Avdeyev, Dylan J. Taylor, et al. 2022. "A Complete Reference Genome Improves Analysis of Human Genetic Variation." *Science* 376 (6588). https://doi.org/10.1126/science.abl3533.
3. Aibar, Sara, Celia Fontanillo, Conrad Droste, and Javier De Las Rivas. 2015. "Functional Gene Networks: R/Bioc Package to Generate and Analyse Gene Networks Derived from Functional Enrichment and Clustering." *Bioinformatics* 31 (10): 1686–8. https://doi.org/10.1093/bioinformatics/btu864.
4. Argelaguet, Ricard, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C. Marioni, and Oliver Stegle. 2020. "MOFA+: A Statistical Framework for Comprehensive Integration of Multi-Modal Single-Cell Data." *Genome Biology* 21 (1). https://doi.org/10.1186/s13059-020-02015-1.
5. Aryee, Martin J., Andrew E. Jaffe, Hector Corrada-Bravo, Christine Ladd-Acosta, Andrew P. Feinberg, Kasper D. Hansen, and Rafael A. Irizarry. 2014. "Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium Dna Methylation Microarrays." *Bioinformatics* 30 (10): 1363–9. https://doi.org/10.1093/bioinformatics/btu049.

6. Cavalcante, Raymond G, and Maureen A Sartor. 2017. "Annotatr: Genomic Regions in Context." Edited by Alfonso Valencia. *Bioinformatics* 33 (15): 2381–3. https://doi.org/10.1093/bioinformatics/btx183.
7. Couto, Bárbara Zita Peters, Nicholas Robertson, Ellis Patrick, and Shila Ghazanfar. 2023. "MoleculeExperiment Enables Consistent Infrastructure for Molecule-Resolved Spatial Transcriptomics Data in Bioconductor." *bioRxiv*. https://doi.org/10.1101/2023.05.16.541040.
8. Gel, Bernat, Anna Diez-Villanueva, Eduard Serra, Marcus Buschbeck, Miguel A. Peinado, and Roberto Malinverni. 2015. "regioneR: An R/Bioconductor Package for the Association Analysis of Genomic Regions Based on Permutation Tests." *Bioinformatics*, September, btv562. https://doi.org/10.1093/bioinformatics/btv562.
9. Hansen, Kasper D., Rafael a. Irizarry, and Zhijin Wu. 2012. "Removing Technical Variability in Rna-Seq Data Using Conditional Quantile Normalization." *Biostatistics* 13: 204–16. https://doi.org/10.1093/biostatistics/kxr054.
10. "Immune Checkpoint Inhibitors." 2022. https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/checkpoint-inhibitors.
11. Kupkova, Kristyna, Jose Verdezoto Mosquera, Jason P. Smith, Micha l Stolarczyk, Tessa L. Danehy, John T. Lawson, Bingjie Xue, John T. Stubbs, Nathan LeRoy, and Nathan C. Sheffield. 2022. "GenomicDistributions: Fast Analysis of Genomic Intervals with Bioconductor." *BMC Genomics* 23 (1). https://doi.org/10.1186/s12864-022-08467-y.
12. Lapuente-Santana, Óscar, Maisa van Genderen, Peter A. J. Hilbers, Francesca Finotello, and Federica Eduati. 2021. "Interpretable Systems Biomarkers Predict Response to Immune-Checkpoint Inhibitors." *Patterns* 2 (8). https://doi.org/10.1016/j.patter.2021.100293.
13. Lawson, John, Eleni Tomazou, Christoph Bock, and Nathan C. Sheffield. 2018. "MIRA: An R Package for DNA Methylation-Based Inference of Regulatory Activity." *Bioinformatics* bty083 (March). https://doi.org/10.1093/bioinformatics/bty083.
14. Lawson, John T., Jason P. Smith, Stefan Bekiranov, Francine E. Garrett-Bakelman, and Nathan C. Sheffield. 2020. "COCOA: Coordinate Covariation Analysis of Epigenetic Heterogeneity." *Genome Biology* 21 (1). https://doi.org/10.1186/s13059-020-02139-4.
15. Mariathasan, Sanjeev, Shannon J. Turley, Dorothee Nickles, Alessandra Castiglioni, Kobe Yuen, Yulei Wang, Edward E. Kadel III, et al. 2018. "TGFB Attenuates Tumour Response to Pd-L1 Blockade by Contributing to Exclusion of T Cells." *Nature* 554 (7693): 544–48. https://doi.org/10.1038/nature25501.
16. Moses, Lambda, Pétur Helgi Einarsson, Kayla Jackson, Laura Luebbert, A Sina Booeshaghi, Sindri Antonsson, Nicolas Bray, Páll Melsted, and Lior Pachter. 2023. "Voyager: Exploratory Single-Cell Genomics Data Analysis with Geospatial Statistics." *bioRxiv*.
17. Müller, Fabian, Michael Scherer, Yassen Assenov, Pavlo Lutsik, Jörn Walter, Thomas Lengauer, and Christoph Bock. 2019. "RnBeads 2.0: Comprehensive Analysis of Dna Methylation Data." *Genome Biology* 20 (1). https://doi.org/10.1186/s13059-019-1664-9.
18. Ou, Jianhong, Haibo Liu, Jun Yu, Michelle A. Kelliher, Lucio H. Castilla, Nathan D. Lawson, and Lihua Julie Zhu. 2018. "ATACseqQC: A Bioconductor Package for Post-Alignment Quality Assessment of Atac-Seq Data." *BMC Genomics* 19 (1). https://doi.org/10.1186/s12864-018-4559-3.
19. Righelli, Dario, Lukas M Weber, Helena L Crowell, Brenda Pardo, Leonardo Collado-Torres, Shila Ghazanfar, Aaron TL Lun, Stephanie C Hicks, and Davide Risso. 2022. "SpatialExperiment: Infrastructure for Spatially-Resolved Transcriptomics Data in R Using Bioconductor." *Bioinformatics* 38 (11): 3128–31.
20. Schatz, Michael C., Anthony A. Philippakis, Enis Afgan, Eric Banks, Vincent J. Carey, Robert J. Carroll, Alessandro Culotti, et al. 2022. "Inverting the Model of Genomics Data Sharing with the Nhgri Genomic Data Science Analysis, Visualization, and Informatics Lab-Space." *Cell Genomics* 2 (1): 100085. https://doi.org/10.1016/j.xgen.2021.100085.
21. Sheffield, Nathan C., and Christoph Bock. 2016. "LOLA: Enrichment Analysis for Genomic Region Sets and Regulatory Elements in R and Bioconductor." *Bioinformatics* 32 (4): 587–89. https://doi.org/10.1093/bioinformatics/btv612.
22. Silva, Tiago C., Simon G. Coetzee, Nicole Gull, Lijing Yao, Dennis J. Hazelett, Houtan Noushmehr, De-Chen Lin, and Benjamin P. Berman. 2018. "ELMER V.2: An

R/Bioconductor Package to Reconstruct Gene Regulatory Networks from Dna Methylation and Transcriptome Profiles." Edited by Oliver Stegle. *Bioinformatics* 35 (11): 1974–7. https://doi.org/10.1093/bioinformatics/bty902.

23. Stark, Rory, and Gordon Brown. 2011. "DiffBind: Differential Binding Analysis of Chip-Seq Peak Data."
24. Tian, Luyi, Shian Su, Xueyi Dong, Daniela Amann-Zalcenstein, Christine Biben, Azadeh Seidi, Douglas J. Hilton, Shalin H. Naik, and Matthew E. Ritchie. 2018. "ScPipe: A Flexible R/Bioconductor Preprocessing Pipeline for Single-Cell Rna-Sequencing Data." Edited by Mihaela Pertea. *PLOS Computational Biology* 14 (8): e1006361. https://doi.org/10.1371/journal.pcbi.1006361.
25. Weber, Lukas M, Arkajyoti Saha, Abhirup Datta, Kasper D Hansen, and Stephanie C Hicks. 2023. "NnSVG for the Scalable Identification of Spatially Variable Genes Using Nearest-Neighbor Gaussian Processes." *Nature Communications* 14 (1): 4059.
26. Wei, Zheng, Wei Zhang, Huan Fang, Yanda Li, and Xiaowo Wang. 2018. "EsATAC: An Easy-to-Use Systematic Pipeline for Atac-Seq Data Analysis." *Bioinformatics (Oxford, England)*, March. https://doi.org/10.1093/bioinformatics/bty141.
27. Zheng, Shijie C., Charles E. Breeze, Stephan Beck, and Andrew E. Teschendorff. 2018. "Identification of Differentially Methylated Cell Types in Epigenome-Wide Association Studies." *Nature Methods* 15 (12): 1059–66. https://doi.org/10.1038/s41592-018-0213-x.
28. Zhu, Lihua J, Claude Gazin, Nathan D Lawson, Hervé Pagès, Simon M Lin, David S Lapointe, and Michael R Green. 2010. "ChIPpeakAnno: A Bioconductor Package to Annotate ChIP-Seq and ChIP-Chip Data." *BMC Bioinformatics* 11 (1). https://doi.org/10.1186/1471-2105-11-237.